DOCUMENT RESUME

ED 357 032 TM 019 725

AUTHOR Finch, Fredrick; Foertsch, Mary
TITLE Varieties of Performance Assessments.
PUB DATE Jan 93
NOTE 6p.; Paper presented at the Annual Meeting of the Michigan Educational Research Association (January 21, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Educational Assessment; Elementary Secondary Education; *Evaluation Methods; Multiple Choice Tests; *Simulation; Situational Tests; *Student Evaluation; Test Construction; Test Format; Test Selection; Test Use
IDENTIFIERS Alternative Assessment; *Authentic Assessment; *Performance Based Evaluation

ABSTRACT

Performance assessment is reviewed as an emerging form of alternative assessment, focusing on how it has been defined in the research literature, the criteria for evaluating its authenticity, the measurement of process and product, and the link between assessment and instruction. Three important dimensions that must be considered in describing performance tests are the extent to which test is authentic (simulating real life), what is actually being evaluated, and the basic response format of the task. In evaluating authenticity, it is necessary to consider structure and design, grading and scoring, and fairness and equity. Performance tests can focus primarily on the product or on the process, and it is important to recognize where the focus lies. The basic types of response situations include oral, written, and graphic representation tasks. Good instructional activities may inform the design of good assessment tasks, but it cannot be assumed that authentic assessment will automatically result in classroom activities that are more conducive to learning. In practice, a compromise between multiple-choice tests and full-blown simulations of actual situations can be found, even though this may not be the highest form of performance assessment. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
FREDRICK FINCH
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED357032

VARIETIES OF PERFORMANCE ASSESSMENTS

Michigan Educational Research Association
January 21, 1993

Fredrick Finch and Mary Foertsch
The Riverside Publishing Company

The most persistent question in assessment today is whether, and how, to move away from traditional multiple-choice testing toward alternative forms of assessment, such as performance assessment. The enthusiasm of performance assessment advocates, however, has been tempered by the concerns of researchers and educators regarding issues of reliability and cost. At least some of the debate between advocates and critics seems to be caused by a lack of a common definition of performance assessment. A possible reason for this situation is that definitions of performance assessment are still evolving.

The purpose of this paper is take a close look at performance assessment as an emerging form of alternative assessment. Our discussion will focus on how performance assessment has been defined in the research literature, the criteria for evaluating the authenticity of performance assessments, the measurement of process and product, and the link between assessment tasks and instructional activities.

Before attempting to define performance tests, it will be useful to make a few subtle but necessary distinctions between testing and assessment. According to the American Psychological Association (APA), the American Educational Research Association (AERA), and the National Council on Measurement in Education (NCME), a test "may be thought of as a set of tasks or questions intended to elicit particular types of behaviors when presented under standardized conditions and to yield scores that have desirable psychometric properties..." (1974, p. 2). According to Anastasi (1988), "Standardization implies uniformity of procedure in administering and scoring the test. If the scores obtained by different persons are to be comparable, testing conditions must obviously be the same for all" (p. 25).

Although testing and assessment are often used interchangeably, they are not synonymous. In educational settings, assessment refers to a multifaceted process that takes into consideration students' performance on a variety of tasks in a variety of settings or contexts. Assessment also involves an evaluative, interpretive appraisal of performance (Salvia and Ysseldyke, 1978). Simply stated, a test is often a collection of similar exercises measuring a wide (or narrow) range of skills. An assessment uses a variety of methods to assess student capability. But, in order to provide consistent (reliable) and accurate (valid) information, both must be administered using standardized procedures.

What is Performance Assessment?

A quick review of the research literature reveals that there are many definitions of the term "performance assessment." Almost anything that is not a multiple-choice paper and pencil test can, and has, been considered a performance assessment (Frechtling, 1991). Included are short-answer and essay questions, portfolios, research projects, simulations, and dramatizations, among others.

Adkins (1974) defined the term "performance test" to mean a test that is used only to evaluate the manipulation of instruments, physical movements, manual dexterity, and so on. Excluded from Adkins's definition would be any paper-and-pencil test even if it is used for the observation of behavior. On the other hand, intelligence scales such as the Wechsler Intelligence Scale for Children (WISC), in which items require examinees to trace a figure with a pencil, are categorized by some psychologists as performance measures (Kojima, 1990).

TM019725

BEST COPY AVAILABLE

A recent statement by Mehrens (1992) reflects the ambiguity that surrounds the use of the term "performance test":

> Typically what users of the term mean is that the assessment will require the examinee to construct an original response. Some people seem to call short-answer questions or fill-in-the-blank questions performance assessments. However, it is more common in performance assessment for the examiner to observe the process of the construction. (p. 3)

In practice, this "observation" is often based on a scorer's evaluation of a written document produced by a student.

Over 20 years ago, Fitzpatrick and Morrison (1971) stated, "There is no absolute distinction between performance tests and other classes of tests--the performance test is one that is relatively realistic." (p. 238). They also suggest that "performance and product evaluation" is a more complete term that can be used interchangeably with the term "performance test." The description of an assessment as performance and product evaluation takes into account the possibility that the situations or contexts in which performance assessments may be cast vary greatly.

In actual practice, the term "educational performance test" signifies a test for which students are given the opportunity to provide constructed responses rather than select answers from a set of predetermined choices. For example, a true/false item would not be appropriate in a performance test unless the student is asked to explain or justify the choice and the student's score is based on the explanation. If the test score is based solely on whether a true or false choice was selected, the item would be more appropriate for a multiple-choice test than a performance test (Finch, 1991).

Distinguishing Types of Performance Tests

There are at least three important dimensions that must be considered when attempting to describe different types of performance tests. The first dimension concerns the extent to which the test is authentic. That is, how closely do the assessment tasks and the situational context in which they are to be performed simulate "real life"? The second dimension involves what is being evaluated: is the intent to evaluate process or product or both? The third dimension concerns the basic response format required by the task: What form is the response expected to take? All of these dimensions interact within a test and influence how a test is ultimately characterized.

Determining the "Fidelity" of Performance Tests

Performance assessments tend to involve special challenges and require decisions and procedures not usually required for conventional tests. The distinction lies in the degree to which the assessment simulates a "real life" situation. In other words, how authentic is the test? Three general areas should be considered in gauging the authenticity of a performance test.

1. Structure and Design. Performance tests (PTs) are constructed to point the student toward more sophisticated and effective ways to use knowledge. PTs are contextualized complex intellectual challenges, not fragments or static bits or tasks. PTs culminate in the student's own product, for which "content" is to be mastered as a means, not as an end. PTs should assess student habits, strategies, and repertoires. PTs are not simply restricted to recall or recognition; they do not reflect lucky or unlucky one-shot responses. PTs should assess "realistic" complexity, stressing depth more than breadth. In doing so, PTs

necessarily involve somewhat ambiguous tasks or problems, and therefore make student judgment central in approaching, clarifying, and tackling problems.

2. Grading or Scoring Tests. PTs measure what is essential, which is not easily counted. Thus, the criteria for scoring must be relatively complex in order to accommodate the multifaceted responses that students may produce.

3. Fairness and Equity. PTs allow students to show what they can do rather than simply rely on right/wrong answers. There must be room for choice and style in tasks, topics, and methodologies.

Performance Versus Product

What is being evaluated? Performance, product, or both? Some performances have no product or the product is indistinguishable from the process, such as public speaking and dancing. In some cases there are many acceptable variations in the process, making the product alone the focus of the assessment experience. Criteria for distinguishing between the evaluation of performance or product have been suggested by Fitzpatrick and Morrison (1971):

Performance Tests that focus primarily on the process

- are based on a procedure that has clearly defined steps
- make it possible to document the extent to which someone deviates from appropriate procedures
- provide much or all of the evidence needed to evaluate the performance in the way that the performance is carried out.

Performance Tests that focus primarily on the product

- result in a product that can be measured accurately and objectively
- result in a product that contains clear evidence of achievement
- contain a sequence of steps to be followed that is indeterminate or has not been taught

Basic Types of Response Situations

The following classifications represent an attempt to describe the basic characteristics of performance test items in terms of what students are asked to do. We recognize that most performance tests are conglomerates of several different performance situations. That is, students may be asked to perform a variety of different tasks in order to complete a particular performance test. A description of basic response formats follows:

Oral Task. The learner reads aloud while his/her oral reading skills are evaluated; includes providing oral answers to interview or test questions. This category also includes such activities as giving speeches, debating, explaining, and spelling.

Written Task. Short answer, justification, essay, sentence completion, making lists, report writing, letter writing, fill-in-the-blank. Includes "thought experiments" in which the performance requirements are clearly stated but the nature of the response is completely up to the student.

Graphic Representation. Includes drawings, graphs, and charts. Products may be purely symbolic in nature or incorporate written language.

Performance or Demonstration. The learner is evaluated while performing in either contrived or artificial situations. Includes such diverse activities as laboratory experiments, dancing, driving, and dramatizations.

The different activities within each type of task can be placed along a continuum which Fitzpatrick and Morrison (1971) call "fidelity of simulation" (p. 239). For example, the context in which written tasks are performed range from reading a passage about a specific topic to attempting to recreate, within the pages of a test book, the environment in which the task is usually performed. It is also the case that each of these basic response formats may result in an evaluation of process or product, or both. And it is certainly possible that within a performance assessment, several different types of tasks might be required.

Assessment Tasks and Instructional Activities

It is widely acknowledged that performance assessments and instructional activities have a great deal in common (Baron, 1991, Baron et. al, 1989; Marzano et. al., 1989). Both provide opportunities for demonstrating complex thinking skills and strategies. Both provide students with feedback. However, Baron (1991) and Marzano et. al. (1989) have suggested these important differences:

1. Assessment tasks must include scoring criteria, whereas instructional tasks do not.

2. Assessment tasks are viewed as culminations of a series of instructional activities, whereas instructional tasks are designed to develop new skills or learnings. Assessment tasks are designed specifically to measure learning.

3. The teacher's role in instructional activities is to mediate learning, whereas in assessment tasks the teacher's role is virtually "hands off". Marzano et. al (1989) define the teacher's role in an instructional setting as "catalytic or structuring."

Performance tasks and instructional activities may appear to be very similar on the surface. However, their differences are important. According to Marzano et. al (1989) "The assessment task should be similar enough to the instructional task so that its contextual cues elicit the transfer of appropriate skills and dispositions from the instructional situation to the new one" (p. 141). If assessment tasks are too difficult, students will not be able to recognize them as providing opportunities to demonstrate the skills and strategies that have been learned. Also, students and teachers may be inappropriately informed as to students' strengths and weaknesses (Nitko, 1989).

Good instructional activities may inform the design of good assessment tasks. The point has been made that assessment, like instruction, provides an occasion for learning (Wiggins, 1989; Wolf, Bixby, Glenn, and Gardner, 1991). However, as Linn, Baker, and Dunbar (1991) have cautioned, "It cannot be assumed that a more 'authentic' assessment will result in classroom activities that are more conducive to learning. We should not be satisfied, for example, if the introduction of a direct writing assessment led to great amounts of time being devoted to the preparation of brief compositions following a formula that works well in producing highly rated essays in a 20-minute time limit" (p. 17).

In summary, most educational performance assessments (or tests) currently available take the form of paper-and-pencil exercises which require students to provide written responses which are evaluated in terms of both process and product. This reflects a need for relatively inexpensive standardized instruments. There is possibly a reasonable compromise between multiple-choice tests and full-blown simulations of actual situations but they should not be considered to be the highest form of performance assessment. "Essay tests" only move a short way beyond multiple-choice tests on the continuum of fidelity of simulation.

BEST COPY AVAILABLE

References

Adkins, D. C. (1974). *Construction and Analysis of Achievement Tests: The Development of Written and Performance Tests of Achievement for Predicting Job Performance of Public Personnel. United States Civil Service Commission, United States Government Printing Office*. Washington, DC.

AERA, APA, & NCME. (1985). *Standards for Educational and Psychological Testing*. Washington, DC: American Psychological Association.

Anastasi, A. (1988). *Psychological Testing (Sixth Edition)*. New York: Macmillan.

Baron, J. (1991). "Strategies for developing effective performance exercises," *Applied Measurement in Education, 4*, pp. 305-318.

Baron, J., Forgione, P., Rindone, D., Kruglanski, H., & Davey, B. (1989, April). *Toward a new generation of student outcome measures: Connecticut Common Core of Language Assessment*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Finch, F.L. (1991). "Issues in educational performance evaluation. In F. Finch, (Ed.), *Educational Performance Assessment (pp. 1-18)*. Chicago, IL: Riverside Publishing Company.

Fitzpatrick, R., & Morrison, E. (1971). "Performance and product evaluation." In R.L. Thorndike (Ed.), *Educational Measurement (second edition)*. Washington, DC: American Council on Education.

Fretchling, J.A. (1991, Winter). "Performance assessment: Moonstruck or the real thing?" *Educational Measurement, 10*, pp. 23-25.

Kojima, S. (1990). "Performance tests." In Walberg, H.J. & Haertel, G. (Eds.), *The International Encyclopedia of Evaluation (pp. 475-481)*. New York: Pergammon Press.

Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991, November). "Complex, performance based assessment: expectations and validation criteria," *Educational Researcher, 20*, pp. 15-21.

Marzano, R., Brandt, R., Hughes, C.S., Jones, B.F., Presseisen, B.Z., Rankin, S.C., & Suhor, C. (1988). *Dimensions of Thinking*. Alexandria, VA: ASCD.

Mehrens, W. (1992, Spring). "Using performance assessments for accountability purposes. *Educational Measurement, 20*, pp. 3-9.

Nitko, A.J. (1989). "Describing tests that integrated with instruction." In R. Linn (Ed), *Educational Measurement (3rd ed.) (pp. 447-474)*. New York: Macmillan.

Salvia, J., & Ysseldyke, J. (1978). *Assessment In Special and Remedial Education*. Boston, MA: Houghton Mifflin.

Wiggins, G. (1989). "A true test: Toward more naturalistic and equitable assessment." *Phi Delta Kappan, 70*, 703-713.

Wolf, D., Bixby, J., Glenn, J. & Gardner, H. (1991). "To use their minds well: Investigations of new forms of student assessment". In G. Grant (Ed.), *Review of Research in Education (pp. 31-74)*. Washington, DC: American Educational Research Association.